NUCLEIC ACID SEQUENCES FROM *DROSOPHILA MELANOGASTER* THAT ENCODE PROTEINS ESSENTIAL FOR VIABILITY AND USES THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application Serial No. 60/436,442, filed December 23, 2002, which is hereby incorporated by reference in its entirety.

The Sequence Listing associated with the instant disclosure has been submitted as an about 298 kb file on CD-R (in duplicate) instead of on paper. Each CD-R is marked in indelible ink to identify the Applicants, Title, File Name (70201USNP.ST25.txt), Creation Date (November 17, 2003), Computer System (IBM-PC/MS-DOS/MS-Windows), and Docket No. (70201USNP). The Sequence Listing submitted on CD-R is hereby incorporated by reference into the instant disclosure.

FIELD OF INVENTION

The present invention pertains to nucleic acid sequences isolated from *Drosophila melanogaster* that encode proteins essential for viability. The invention particularly relates to methods of using these proteins as insecticide targets, based on this essentiality.

BACKGROUND OF THE INVENTION

Insects contribute or cause many human and animal diseases, and are responsible for substantial agricultural and property damage. The societal costs associated with insect pests in dollars, time and suffering are monumental. The total worldwide market size for insecticide crop protection is over $5 billion. To combat these problems, insecticidal compounds have been developed and employed.

The idea to use chemicals for insect control is not new. The scientific use of pesticides started with the introduction of arsenical insecticides and organic compounds such as tar, petroleum oils, and dinitrophenol emulsions at the end of the last century. But, the systematic search for synthetic organic insecticides was only launched after the discovery of the insecticidal properties of DDT in 1939. After World War II, chemical research concentrated mainly on chlorinated hydrocarbons and cyclodienes, which all require high rates of application and have a rather broad spectrum of activity. Most of them are persistent in the environment and may pose a

significant risk for accumulation in the food chain. Today the use of these chemicals is very much restricted.

From this point, the major emphasis in research has been given to organophosphates and carbamates, which are readily degradable in the environment with little tendency for bioaccumulation. The toxicity of these compounds varies within a broad range from medium to highly toxic. Organophosphates and carbamates are still widely use, although the more toxic ones are banned in certain countries. The formamidines have as their major advantage a different mode of action and their selectivity, which made them suitable for use in IPM (insect pest management) programs. They are easily degradable with no accumulation potential, but for toxicological reasons some have had to be withdrawn from the market.

For the past decade, insecticide research has concentrated on leadfinding for new chemical structures interfering with new target mechanisms. The chances for success are rather remote, because the hurdles for the registration of a new insecticide are set very high. Toxicological aspects, insecticide resistance, environmental behavior, and IPM fitness are some of the critical factors that have to be considered together with economical factors.

Novel insecticides can now be discovered using high-throughput screens that implement recombinant DNA technology. Proteins found to be essential to insect viability can be recombinantly produced through standard molecular biological techniques and utilized as insecticide targets in screens for novel inhibitors of the enzymes' activity. The novel inhibitors discovered through such screens may then be used as insecticides to control undesirable insect infestation.

However, as the world population continues to grow, there will be increasing food shortages. Therefore, there exists continuing need to find new, effective and economic insecticides.

SUMMARY OF THE INVENTION

In view of these needs, it is one object of the invention to provide essential genes in insects such as *Drosophila melanogaster*. It is another object to provide the essential proteins encoded by these essential genes for assay development to identify inhibitory compounds with insecticidal activity. It is still another object of the present invention to provide an effective and beneficial method for identifying new or improved insecticides using the essential proteins of the invention.

In furtherance of these and other objects, the present invention provides DNA molecules comprising nucleotide sequences isolated from *Drosophila melanogaster* that encode proteins essential for viability. The inventors are the first to demonstrate that the nucleotide sequences of the invention are essential for viability. This knowledge is exploited to provide novel insecticide modes of action. One advantage of the present invention is that the proteins encoded by the essential nucleotide sequences provide the bases for assays designed to easily and rapidly identify novel insecticides.

Disruption of the nucleotide sequences or messenger RNA of the invention demonstrates that the activity of each corresponding encoded protein is essential for *Drosophila* viability. Genetic results show that when each nucleotide sequence of the invention is mutated in *Drosophila* or disrupted at the transcription level, the resulting phenotype is lethal.. This demonstrates a critical role for the protein encoded by the mutated nucleotide sequence. This further implies that chemicals that inhibit the expression of the protein when in contact with insects are likely to have detrimental effects on insects and are potentially good insecticide candidates. The present invention therefore provides methods of using the disclosed nucleotide sequences or proteins encoded thereby to identify inhibitors thereof. The inhibitors can then be used as insecticides to kill undesirable insect populations where crops are grown, particularly agronomically important crops such as maize, and other cereal crops such as wheat, oats, rye, sorgum, rice, barley, millet, turf and forage grasses and the like, as well as cotton, sugar cane, sugar beet, oilseed rape, soybeans, vegetable crops and fruits.

The present invention accordingly provides cDNA sequences derived from *Drosophila melanogaster*. In one embodiment, the present invention provides an isolated DNA molecule comprising a nucleotide sequence selected from the group consisting of the odd numbered SEQ ID NOs:1-49. In another embodiment, the present invention provides an isolated DNA molecule comprising a nucleotide sequence that encodes a protein selected from the group consisting of the even numbered SEQ ID NOs:2-50.

The present invention also provides a chimeric construct comprising a promoter operatively linked to a DNA molecule according to the present invention, wherein the promoter is preferably functional in a eukaryote, wherein the promoter is preferably heterologous to the DNA molecule. The present invention further provides a recombinant vector comprising a chimeric construct

according to the present invention, wherein said vector is capable of being stably transformed into a host cell. The present invention still further provides a host cell comprising a DNA molecule according to the present invention, wherein said DNA molecule is preferably expressible in the cell. The host cell is preferably selected from the group consisting of an insect cell, a yeast cell, and a prokaryotic cell.

The present invention also provides proteins essential for *Drosophila melanogaster* viability. In one embodiment, the present invention provides an isolated protein comprising an amino acid sequence selected from the group consisting of the even numbered SEQ ID NOs:2-50. In accordance with another embodiment, the present invention also relates to the recombinant production of proteins of the invention and methods of using the proteins of the invention in assays for identifying compounds that interact with the protein.

In another preferred embodiment, the present invention describes a method for identifying chemicals having the ability to inhibit the activity of the disclosed proteins. In a preferred embodiment, the present invention provides a method for selecting compounds that interact with a protein of the invention, comprising: (a) expressing a DNA molecule according to the present invention to generate the corresponding protein of the invention, (b) testing a compound suspected of having the ability to interact with the protein expressed in step (a), and (c) selecting compounds that interact with the protein in step (b).

Other objects and advantages of the present invention will become apparent to those skilled in the art and from a study of the following description of the invention and non-limiting examples. The entire contents of all publications mentioned herein are hereby incorporated by reference.

BRIEF DESCRIPTION OF THE SEQUENCES IN THE SEQUENCE LISTING

Odd numbered SEQ ID NOs:1-49 are nucleotide sequences described in the table below.

Even numbered SEQ ID NOs:2-50 are protein sequences encoded by the immediately preceding nucleotide sequence, e.g., SEQ ID NO:2 is the protein encoded by the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:16 is the protein encoded by the nucleotide sequence of SEQ ID NO:15, etc.

SEQ ID NOs:51-63 are PCR primers.

Table 1 Drosophila Sequences

seq ID	Inventor's reference	function	domains	Best blast hit	score
1-2	GIN00418, CT41283	unknown	IPR001994 == CytidylyltransferaseSCOP:52374 == Nucleotidylyl transferase;	gb|EAA03660.1| agCP14767 [Anopheles gambiae str. PEST]	372
3-4	GIN00831, CT1581	protein serine/threonine kinase	IPR000719 == Eukaryotic protein kinaseSCOP:56112 == Protein kinase-like (PK-like	gb|EAA12395.1| agCP11315 [Anopheles gambiae str. PEST]	555
5-6	GIN00996, CT4870	unknown		gb|EAA00732.1| agCP9812 [Anopheles gambiae str. PEST]	165
7-8	GIN01641, CT4036	chaperone	IPR001580 == Calreticulin family	gb|EAA09483.1| agCP14905 [Anopheles gambiae str. PEST]	637
9-10	GIN02024, CT27956	proton transport	IPR000194 == ATP synthase &agr; and &bgr; subunit, N-terminal"IPR000793 , C-terminal	gb|EAA08458.1| agCP2933 [Anopheles gambiae str. PEST]	957
11-12	GIN05114, CT35627	unknown	IPR003656 == BED finger	gb|EAA15042.1| agCP4573 [Anopheles gambiae str. PEST]	62
13-14	GIN05842, CT4886	V-type ATPase	IPR002490 == V-type ATPase 116kDa subunit family	gb|EAA08852.1| ebiP428 [Anopheles gambiae str. PEST]	1124
15-16	GIN06014, CT32725	protein phosphatase type 2A	SCOP:48371 == ARM repeat;	gb|EAA14749.1| agCP4924 [Anopheles gambiae str. PEST]	842
17-18	GIN08020, CT11825	cyclin-dependent protein kinase	SCOP:47954 == Cyclin-like	gb|EAA10346.1| agCP2112 [Anopheles gambiae str. PEST]	241
19-20	GIN08522, CT13682	translation initiation factor	IPR002735 == Domain found in IF2B/IF5	gb|EAA04210.1| agCP3862 [Anopheles gambiae str. PEST]	285
21-22	GIN08754, CT14494	heat shock protein	SCOP:52821 == Rhodanese/Cell cycle control phosphatase	b|EAA04564.1| agCP3860 [Anopheles gambiae str. PEST]	109
23-24	GIN09345, CT16487	unknown		gb|EAA04319.1| agCP3728 [Anopheles gambiae str. PEST]	81

25-26	GIN09460, CT16853	N-acetylglucosamine-1-phosphate transferase	IPR000715 == Glycosyl transferase, family 4	gb|EAA12410.1| agCP10833 [Anopheles gambiae str. PEST]	401
27-28	GIN09658, CT17430	chaperonin ATPase	SCOP:48592 == GroEL-like chaperones, ATPase domain	gb|EAA05907.1| agCP14562 [Anopheles gambiae str. PEST]	749
29-30	GIN10467, CT20131	unknown	none	none	
31-32	GIN10517, CT20377	protein tyrosine phosphatase	SCOP:52799 == (Phosphotyrosine protein) phosphatases II	dbj|BAA33720.1| EDTP(egg derived tyrosine phosphatase), [Sarcophaga peregrina]	674
33-34	GIN10694, CT20945	nuclear pore protein		gb|EAA00821.1| agCP12701 [Anopheles gambiae str. PEST] 546	546
35-36	GIN10918, CT21672	Vacuolar ATP synthase 16kD subunit	IPR000245 == Vacuolar ATP synthase 16kD subunit	gb|EAA05773.1| ebiP3500 [Anopheles gambiae str. PEST]	223
37-38	GIN11550, CT20832	CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase	IPR001736 == Phospholipase D/Transphosphatidylase	gb|EAA08154.1| agCP1721 [Anopheles gambiae str. PEST]	477
39-40	GIN11578, CT23580	DNA topoisomerase (ATP-hydrolyzing)	IPR002815 == Type II DNA topoisomerase"SCOP:56726 == DNA topoisomerase IV, alpha subunit	gb|AAH33591.1| Similar to SPO11 meiotic protein covalently bound to DSB-like (S. cerevisiae) [Homo sapiens]	149
41-42	GIN11589, CT23419	general RNA polymerase II transcription factor		gb|EAA05440.1| agCP10546 [Anopheles gambiae str. PEST]	644
43-44	GIN11844, CT24166	translation initiation factor	IPR001253 == Eukaryotic initiation factor 1ASCOP:50249 == Nucleic acid-binding proteins	gb|EAA08471.1| agCP2987 [Anopheles gambiae str. PEST]	197
45-46	GIN11932, CT20784	unknown	IPR003006 == Immunoglobulin and major histocompatibility complex domainIPR003598 == Immunoglobulin C-2 type	gb|EAA14754.1| ebiP5214 [Anopheles gambiae str. PEST]	319

47-48	GIN12213, CT24821	ARF small monomeric GTPase	SCOP:48425 == Sec7 domain	dbj|BAA13379.2| Similar to S.cerevisiae SEC7 protein (A31068) [Homo sapiens]	1275
49-50	GIN12858, CT26398	sodium/potassium-exchanging ATPase	IPR000402 == Na+,K+ ATPase &bgr; subunit	gb|EAA12679.1| ebiP2356 [Anopheles gambiae str. PEST]	433

DEFINITIONS

For clarity, certain terms used in the specification are defined and used as follows:

"Associated with / operatively linked" refer to two nucleic acid sequences that are related physically or functionally. For example, a promoter or regulatory DNA sequence is said to be "associated with" a DNA sequence that codes for an RNA or a protein if the two sequences are operatively linked, or situated such that the regulator DNA sequence will affect the expression level of the coding or structural DNA sequence.

A "chimeric construct" is a recombinant nucleic acid sequence in which a promoter or regulatory nucleic acid sequence is operatively linked to, or associated with, a nucleic acid sequence that codes for an mRNA or which is expressed as a protein, such that the regulatory nucleic acid sequence is able to regulate transcription or expression of the associated nucleic acid sequence. The regulatory nucleic acid sequence of the chimeric construct is not normally operatively linked to the associated nucleic acid sequence as found in nature.

Co-factor: natural reactant, such as an organic molecule or a metal ion, required in an enzyme-catalyzed reaction. A co-factor is e.g. NAD(P), riboflavin (including FAD and FMN), folate, molybdopterin, thiamin, biotin, lipoic acid, pantothenic acid and coenzyme A, S-adenosylmethionine, pyridoxal phosphate, ubiquinone, menaquinone. Optionally, a co-factor can be regenerated and reused.

A "coding sequence" is a nucleic acid sequence that is transcribed into RNA such as mRNA, rRNA, tRNA, snRNA, sense RNA or antisense RNA. Preferably the RNA is then translated in an organism to produce a protein.

Complementary: "complementary" refers to two nucleotide sequences that comprise antiparallel nucleotide sequences capable of pairing with one another upon formation of

hydrogen bonds between the complementary base residues in the antiparallel nucleotide sequences.

"Conservatively modified variations" of a particular nucleic acid sequence refers to those nucleic acid sequences that encode identical or essentially identical amino acid sequences, or where the nucleic acid sequence does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given polypeptide. For instance the codons CGT, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded protein. Such nucleic acid variations are "silent variations" which are one species of "conservatively modified variations." Every nucleic acid sequence described herein which encodes a protein also describes every possible silent variation, except where otherwise noted. One of skill will recognize that each codon in a nucleic acid (except ATG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each "silent variation" of a nucleic acid which encodes a protein is implicit in each described sequence.

Furthermore, one of skill will recognize that individual substitutions deletions or additions that alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 1%) in an encoded sequence are "conservatively modified variations," where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following five groups each contain amino acids that are conservative substitutions for one another: Aliphatic: Glycine (G), Alanine (A), Valine (V), Leucine (L), Isoleucine (I); Aromatic: Phenylalanine (F), Tyrosine (Y), Tryptophan (W); Sulfur-containing: Methionine (M), Cysteine (C); Basic: Arginine (R), Lysine (K), Histidine (H); Acidic: Aspartic acid (D), Glutamic acid (E), Asparagine (N), Glutamine (Q). *See also,* Creighton (1984) *Proteins, W.H.* Freeman and Company. In addition, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence are also "conservatively modified variations."

DNA Shuffling: DNA shuffling is a method to rapidly, easily and efficiently introduce mutations or rearrangements, preferably randomly, in a DNA molecule or to generate exchanges of DNA sequences between two or more DNA molecules, preferably randomly. The DNA molecule resulting from DNA shuffling is a shuffled DNA molecule that is a non-naturally occurring DNA molecule derived from at least one template DNA molecule. The shuffled DNA encodes an enzyme modified with respect to the enzyme encoded by the template DNA, and preferably has an altered biological activity with respect to the enzyme encoded by the template DNA.

Enzyme/Protein Activity: means herein the ability of an enzyme (or protein) to catalyze the conversion of a substrate into a product. A substrate for the enzyme comprises the natural substrate of the enzyme but also comprises analogues of the natural substrate, which can also be converted, by the enzyme into a product or into an analogue of a product. The activity of the enzyme is measured for example by determining the amount of product in the reaction after a certain period of time, or by determining the amount of substrate remaining in the reaction mixture after a certain period of time. The activity of the enzyme is also measured by determining the amount of an unused co-factor of the reaction remaining in the reaction mixture after a certain period of time or by determining the amount of used co-factor in the reaction mixture after a certain period of time. The activity of the enzyme is also measured by determining the amount of a donor of free energy or energy-rich molecule (e.g. ATP, phosphoenolpyruvate, acetyl phosphate or phosphocreatine) remaining in the reaction mixture after a certain period of time or by determining the amount of a used donor of free energy or energy-rich molecule (e.g. ADP, pyruvate, acetate or creatine) in the reaction mixture after a certain period of time.

Essential: an "essential" *Drosophila melanogaster* nucleotide sequence is a nucleotide sequence encoding a protein such as e.g. a biosynthetic enzyme, receptor, signal transduction protein, structural gene product, or transport protein that is essential to the growth or survival of the insect.

Expression Cassette: "Expression cassette" as used herein means a DNA sequence capable of directing expression of a particular nucleotide sequence in an appropriate host cell, comprising a promoter operatively linked to the nucleotide sequence of interest which is

operatively linked to termination signals. It also typically comprises sequences required for proper translation of the nucleotide sequence. The coding region usually codes for a protein of interest but may also code for a functional RNA of interest, for example antisense RNA or a nontranslated RNA, in the sense or antisense direction. The expression cassette comprising the nucleotide sequence of interest may be chimeric, meaning that at least one of its components is heterologous with respect to at least one of its other components. The expression cassette may also be one which is naturally occurring but has been obtained in a recombinant form useful for heterologous expression. Typically, however, the expression cassette is heterologous with respect to the host, i.e., the particular DNA sequence of the expression cassette does not occur naturally in the host cell and must have been introduced into the host cell or an ancestor of the host cell by a transformation event. The expression of the nucleotide sequence in the expression cassette may be under the control of a constitutive promoter or of an inducible promoter which initiates transcription only when the host cell is exposed to some particular external stimulus. In the case of a multicellular organism, such as an insect, the promoter can also be specific to a particular tissue or organ or stage of development.

Gene: the term "gene" is used broadly to refer to any segment of DNA associated with a biological function. Thus, genes include coding sequences and/or the regulatory sequences required for their expression. Genes also include nonexpressed DNA segments that, for example, form recognition sequences for other proteins. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

Heterologous/exogenous: The terms "heterologous" and "exogenous" when used herein to refer to a nucleic acid sequence (e.g. a DNA sequence) or a gene, refer to a sequence that originates from a source foreign to the particular host cell or, if from the same source, is modified from its original form. Thus, a heterologous gene in a host cell includes a gene that is endogenous to the particular host cell but has been modified through, for example, the use of DNA shuffling. The terms also include non-naturally occurring multiple copies of a naturally occurring DNA sequence. Thus, the terms refer to a DNA segment that is foreign or heterologous to the cell, or homologous to the cell but in a position within the host cell nucleic acid in which the element is not ordinarily found. Exogenous DNA segments are expressed to

yield exogenous polypeptides.

A "homologous" nucleic acid (e.g. DNA) sequence is a nucleic acid (e.g. DNA) sequence naturally associated with a host cell into which it is introduced.

The terms "identical" or percent "identity" in the context of two or more nucleic acid or protein sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection.

Inhibitor: a chemical substance that inactivates the enzymatic activity of an enzyme (or protein) of interest. The term "insecticide" is used herein to define an inhibitor when applied to an insect at any stage of development.

Insecticide: a chemical substance used to kill or inhibit the growth or viability of insects at any stage of development.

Interaction: quality or state of mutual action such that the effectiveness or toxicity of one protein or compound on another protein is inhibitory (antagonists) or enhancing (agonists).

A nucleic acid sequence is "isocoding with" a reference nucleic acid sequence when the nucleic acid sequence encodes a polypeptide having the same amino acid sequence as the polypeptide encoded by the reference nucleic acid sequence.

An "isolated" nucleic acid molecule or an isolated enzyme is a nucleic acid molecule or enzyme that, by the hand of man, exists apart from its native environment and is therefore not a product of nature. An isolated nucleic acid molecule or enzyme may exist in a purified form or may exist in a non-native environment such as, for example, a recombinant host cell.

Mature Protein: protein that is normally targeted to a cellular organelle and from which the transit peptide has been removed.

Minimal Promoter: promoter elements, particularly a TATA element, that are inactive or that have greatly reduced promoter activity in the absence of upstream activation. In the presence of a suitable transcription factor, the minimal promoter functions to permit transcription.

Modified Enzyme Activity: enzyme activity different from that which naturally occurs in an insect (i.e. enzyme activity that occurs naturally in the absence of direct or indirect manipulation of such activity by man), which is tolerant to inhibitors that inhibit the naturally

occurring enzyme activity.

Native: refers to a gene that is present in the genome of an untransformed insect cell.

Naturally occurring: the term "naturally occurring" is used to describe an object that can be found in nature as distinct from being artificially produced by man. For example, a protein or nucleotide sequence present in an organism (including a virus), which can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory, is naturally occurring.

Nucleic acid: the term "nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides which have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof *(e.g.* degenerate codon substitutions) and complementary sequences and as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al., Nucleic Acid Res.* 19: 5081 (1991); Ohtsuka *et al., J. Biol. Chem.* 260: 2605-2608 (1985); Rossolini *et al., Mol. Cell. Probes* 8: 91-98 (1994)). The terms "nucleic acid" or "nucleic acid sequence" may also be used interchangeably with gene, cDNA, and mRNA encoded by a gene.

"ORF" means open reading frame.

Purified: the term "purified," when applied to a nucleic acid or protein, denotes that the nucleic acid or protein is essentially free of other cellular components with which it is associated in the natural state. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein which is the predominant species present in a preparation is substantially purified. The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or

protein is at least about 50% pure, more preferably at least about 85% pure, and most preferably at least about 99% pure.

Two nucleic acids are "recombined" when sequences from each of the two nucleic acids are combined in a progeny nucleic acid. Two sequences are "directly" recombined when both of the nucleic acids are substrates for recombination. Two sequences are "indirectly recombined" when the sequences are recombined using an intermediate such as a cross-over oligonucleotide. For indirect recombination, no more than one of the sequences is an actual substrate for recombination, and in some cases, neither sequence is a substrate for recombination.

"Regulatory elements" refer to sequences involved in controlling the expression of a nucleotide sequence. Regulatory elements comprise a promoter operatively linked to the nucleotide sequence of interest and termination signals. They also typically encompass sequences required for proper translation of the nucleotide sequence.

Significant Increase: an increase in enzymatic activity that is larger than the margin of error inherent in the measurement technique, preferably an increase by about 2-fold or greater of the activity of the wild-type enzyme in the presence of the inhibitor, more preferably an increase by about 5-fold or greater, and most preferably an increase by about 10-fold or greater.

Substantially identical: the phrase "substantially identical," in the context of two nucleic acid or protein sequences, refers to two or more sequences or subsequences that have at least 60%, preferably 80%, more preferably 90, even more preferably 95%, and most preferably at least 99% nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. Preferably, the substantial identity exists over a region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably the sequences are substantially identical over at least about 150 residues. In an especially preferred embodiment, the sequences are substantially identical over the entire length of the coding regions. Furthermore, substantially identical nucleic acid or protein sequences perform substantially the same function.

For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated if necessary, and

sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2: 482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48: 443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85: 2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (*see generally,* Ausubel *et al., infra).*

One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al., J. Mol. Biol.* 215: 403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information on the world wide web at ncbi.nlm.nih.gov/. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, 1990). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino

acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix *(see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89: 10915 (1989)).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences *(see, e.g.,* Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90: 5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid sequence to the reference nucleic acid sequence is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions. The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (*e.g.,* total cellular) DNA or RNA. "Bind(s) substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target nucleic acid sequence.

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern hybridizations are sequence dependent, and are different under different environmental parameters. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic* Acid Probes part I chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays" Elsevier, New York. Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH.

Typically, under "stringent conditions" a probe will hybridize to its target subsequence, but to no other sequences.

The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formamide with 1 mg of heparin at 42°C, with the hybridization being carried out overnight. An example of highly stringent wash conditions is 0.1 5M NaCl at 72°C for about 15 minutes. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see,* Sambrook, *infra*, for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1x SSC at 45°C for 15 minutes. An example low stringency wash for a duplex of, *e.g.,* more than 100 nucleotides, is 4-6x SSC at 40°C for 15 minutes. For short probes (*e.g.,* about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1.0 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is typically at least about 30°C. Stringent conditions can also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the proteins that they encode are substantially identical. This occurs, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

The following are examples of sets of hybridization/wash conditions that may be used to clone homologous nucleotide sequences that are substantially identical to reference nucleotide sequences of the present invention: a reference nucleotide sequence preferably hybridizes to the reference nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 2X SSC, 0.1% SDS at 50°C, more desirably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS at 50°C, more desirably still in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C

with washing in 0.5X SSC, 0.1% SDS at 50°C, preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 50°C, more preferably in 7% sodium dodecyl sulfate (SDS), 0.5 M $NaPO_4$, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C.

A further indication that two nucleic acid sequences or proteins are substantially identical is that the protein encoded by the first nucleic acid is immunologically cross reactive with, or specifically binds to, the protein encoded by the second nucleic acid. Thus, a protein is typically substantially identical to a second protein, for example, where the two proteins differ only by conservative substitutions.

The phrase "specifically (or selectively) binds to an antibody," or "specifically (or selectively) immunoreactive with," when referring to a protein or peptide, refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, antibodies raised to the protein with the amino acid sequence encoded by any of the nucleic acid sequences of the invention can be selected to obtain antibodies specifically immunoreactive with that protein and not with other proteins except for polymorphic variants. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays, Western blots, or immunohistochemistry are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane (1988) *Antibodies, A Laboratory Manual,* Cold Spring Harbor Publications, New York "Harlow and Lane"), for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity. Typically a specific or selective reaction will be at least twice background signal or noise and more typically more than 10 to 100 times background.

A "subsequence" refers to a sequence of nucleic acids or amino acids that comprise a part of a longer sequence of nucleic acids or amino acids (e.g., protein) respectively.

"Synthetic" refers to a nucleotide sequence comprising structural characters that are not present in the natural sequence. For example, an artificial sequence that resembles more closely the G+C content and the normal codon distribution of dicot and/or monocot genes is said to be synthetic.

Substrate: a substrate is the molecule that an enzyme naturally recognizes and converts to a product in the biochemical pathway in which the enzyme naturally carries out its function, or is a modified version of the molecule, which is also recognized by the enzyme and is converted by the enzyme to a product in an enzymatic reaction similar to the naturally-occurring reaction.

Target gene: A "target gene" is any gene in an insect cell. For example, a target gene is a gene of known function or is a gene whose function is unknown, but whose total or partial nucleotide sequence is known. Alternatively, the function of a target gene and its nucleotide sequence are both unknown. A target gene is a native gene of the insect cell or is a heterologous gene that had previously been introduced into the insect cell or a parent cell of said insect cell, for example by genetic transformation. A heterologous target gene is stably integrated in the genome of the insect cell or is present in the insect cell as an extrachromosomal molecule, e.g. as an autonomously replicating extrachromosomal molecule.

Transformation: a process for introducing heterologous DNA into a cell, tissue, or insect. Transformed cells, tissues, or insects are understood to encompass not only the end product of a transformation process, but also transgenic progeny thereof.

"Transformed," "transgenic," and "recombinant" refer to a host organism such as a bacterium or a plant into which a heterologous nucleic acid molecule has been introduced. The nucleic acid molecule can be stably integrated into the genome of the host or the nucleic acid molecule can also be present as an extrachromosomal molecule. Such an extrachromosomal molecule can be auto-replicating. Transformed cells, tissues, or plants are understood to encompass not only the end product of a transformation process, but also transgenic progeny thereof. A "non-transformed," "non-transgenic," or "non-recombinant" host refers to a wild-type organism, e.g., a bacterium or plant, which does not contain the heterologous nucleic acid molecule.

Viability: "viability" as used herein refers to a fitness parameter of an insect. Insects are assayed for their homozygous performance of *Drosophila* development, indicating which proteins are indispensable to maintain life in *Drosophila*.

DETAILED DESCRIPTION OF THE INVENTION

I. Identification Of Essential *Drosophila melanogaster* Nucleotide Sequences Using Transposable Element Insertion Mutagenesis

As shown in Table 2 and the examples below, the identification of novel nucleotide sequences, as well as the essentiality of the nucleotide sequences for normal insect viability, have been demonstrated in *Drosophila* using P-element transposable insertion mutagenesis. Having established the essentiality of the function of the encoded proteins in *Drosophila* and having identified the nucleotide sequences encoding these essential proteins, the inventors thereby provide an important and sought-after tool for new insecticide development.

A lethal phenotype caused by insertion of a P-element indicates that the affected nucleotide sequence codes for an essential protein in the insect. The characterization of the insertion site using flanking sequence DNA is needed to associate an individual lethal line with specific nucleotide sequences. Genomic DNA adjacent to the 5' and/or 3' end of the P-element from the insertion line is generated using inverse PCR.

Table 2 Method of validation of nucleic acid sequences as essential

seq ID	Inventor's reference	in house validation method
1	GIN00418, CT41283	p-element disruption
3	GIN00831, CT1581	p-element disruption
5	GIN00996, CT4870	p-element disruption
7	GIN01641, CT4036	p-element disruption
9	GIN02024, CT27956	dsRNA
11	GIN05114, CT35627	p-element disruption
13	GIN05842, CT4886	dsRNA
15	GIN06014, CT32725	dsRNA
17	GIN08020, CT11825	dsRNA
19	GIN08522, CT13682	p-element disruption
21	GIN08754, CT14494	p-element disruption
23	GIN09345, CT16487	p-element disruption
25	GIN09460, CT16853	dsRNA
27	GIN09658, CT17430	p-element disruption
29	GIN10467, CT20131	dsRNA
31	GIN10517, CT20377	dsRNA

33	GIN10694, CT20945	p-element disruption
35	GIN10918, CT21672	p-element disruption
37	GIN11550, CT20832	dsRNA
39	GIN11578, CT23580	dsRNA
41	GIN11589, CT23419	p-element disruption
43	GIN11844, CT24166	p-element disruption
45	GIN11932, CT20784	dsRNA
47	GIN12213, CT24821	dsRNA
49	GIN12858, CT26398	p-element disruption and dsRNA

I. Determining The Complete Coding Sequences Of The Essential *Drosophila* Nucleotide Sequences

The essential *Drosophila* nucleotide sequences are identified by isolating nucleotide sequences flanking the P-element insertion and aligning that sequence with genomic *Drosophila* sequence obtained from the Celera *Drosophila* database. The protein prediction for each genomic region is obtained by use of an exon algorithm program such as GeneMark. All exon algorithm programs currently used for prediction of proteins are susceptible to inaccuracies, including incomplete predictions of coding sequences, missing alternative splice variants, combining of nearby exons of adjacent genes, and mistranslation at intron-exon borders. The prediction of a complete coding sequence can be confirmed by several methods including polymerase chain reaction (PCR) amplification using the 5' and 3' sequence to verify the message, reverse transcription PCR (rtPCR) using an oligonucleotide internal sequence to identify the 5' and/or 3' end, and screening of cDNA libraries from insect tissues with probes made from a particular sequence to isolate a true full-length clone. To confirm that the message size is accurate, a Northern blot can be hybridized with a probe from the nucleotide sequence. In addition, matches to the *Drosophila* EST database helps to confirm existence of message and gives information about the temporal and spatial pattern of expression. Mutation-causing P elements are known to preferentially cluster in the 5' region of affected genes (Spradling *et al., Proc. Natl. Acad. Sci. USA* 92: 10824-10830 (1995)), a tendency that increases the chance of recovering overlaps between short flanking sequences and 5' ESTs. The present invention therefore provides a number of essential nucleotide sequences as well as the amino acid sequences encoded thereby. cDNA clone sequences are set forth in even numbered SEQ ID

NOs:14-380. The corresponding encoded amino acid sequences are set forth in odd numbered SEQ ID NOs:15-381.

The isolated gene sequences disclosed herein may be manipulated according to standard genetic engineering techniques to suit any desired purpose. For example, an entire *Drosophila* gene sequence or portions thereof may be used as a probe capable of specifically hybridizing to coding sequences and messenger RNAs. To achieve specific hybridization under a variety of conditions, such probes include, e.g. sequences that are unique among insect nucleotide sequences for a particular protein of interest and are at least 10 nucleotides in length, preferably at least 20 nucleotides in length, and most preferably at least 50 nucleotides in length. Such probes are used to amplify and analyze related nucleotide sequences from a chosen organism via PCR. This technique is useful to isolate additional insect nucleotide sequences from a desired organism or as a diagnostic assay to determine the presence of particular nucleotide sequences in an organism. This technique also is used to detect the presence of altered nucleotide sequences associated with a particular condition of interest such as insecticide tolerance, poor health, etc.

Gene-specific hybridization probes also are used to quantify levels of a particular gene mRNA in an insect using standard techniques such as Northern blot analysis. This technique is useful as a diagnostic assay to detect altered levels of gene expression that are associated with particular conditions such as enhanced tolerance to insecticides that target a particular gene.

I.A. Identification of Essential *Drosophila melannogaster* Nucleotide Sequences using RNAi

RNA-mediated interference (RNAi) is a recently discovered method to determine gene function in a number of organisms, wherein double-stranded RNA (dsRNA) directs gene-specific, post-transcriptional silencing. See, e.g., Kuwabara & Olson (2000) Parasitol Today 16(8):347-349; Bass (2000) Cell 101(3):235-238; Hunter (2000) Curr Biol 10(4):R137-140; Bosher & Labouesse (2000) Nat Cell Biol 2(2):E31-36; Sharp (1999) Genes Dev 13(2):139-141. The double-stranded RNA molecule can be synthesized in vitro and then introduced into the organism by injection or other methods. Alternatively, a heritable transgene exhibiting dyad symmetry can provide a transcript that folds as a hairpin structure. Methods for examining gene functions using dsRNAi in Drosophila are disclosed in Example 4a and further in Kennerdell &

Carthew (2000) Nat Biotech 18(8):896-898; Lam & Thummel (2000) Curr Biol 10(16):957-963; Misquitta & Paterson (1999) Proc Natl Acad Sci USA 96 (4):1451-1456.

The present invention describes RNA-mediated interference of sequences listed in Table 2 and Table 6. Double-stranded RNA complementary to each sequence was synthesized in vitro and injected into early Drosophila embryos, as described in Example 4a. Development of injected embryos was assessed by scoring: (a) morphological criteria using a light microscope (Campos-Ortega & Hartenstein (1985) The Embryonic Development of Drosophila melanogaster, Springer-Verlag, Berlin), (b) embryo hatching to become a larvae, (c) puparium formation, and (d) eclosion of the pupae as an adult fly, as indicated in Table 6 herein below. Buffer-injected embryos were injected and monitored in parallel as a control. The percentage of embryos injected with dsRNA that survive to the adult stage is depicted in set forth in Table 6.

Essential genes were identified as those resulting in a percent viable adults below 38% when disrupted by RNAi. This threshold was determined by comparison to multiple buffer-injected controls.

II. Recombinant Production Of Protein And Uses Thereof

For recombinant production of a protein of the invention in a host organism, a nucleotide sequence encoding the protein is inserted into an expression cassette designed for the chosen host and introduced into the host where it is recombinantly produced. The choice of the specific regulatory sequences such as promoter, signal sequence, 5' and 3' untranslated sequence, and enhancer appropriate for the chosen host is within the level of the skill of the routineer in the art. The resultant molecule, containing the individual elements linking in the proper reading frame, is inserted into a vector capable of being transformed into the host cell. Suitable expression vectors and methods for recombinant production of proteins are well known for host organisms such as E. coli, yeast, and insect cells (see, e.g., Lucknow and Summers, Bio/Technol. 6:47 (1988)). Additional suitable expression vectors are baculovirus expression vectors, e.g., those derived from the genome of Autographica californica nuclear polyhedrosis virus (AcMNPV). A preferred baculovirus/insect system is PVL1392(3) used to transfect Spodoptera frugiperda SF9 cells (ATCC) in the presence of linear Autographica californica baculovirus DNA (Phramingen, San Diego, CA). The resulting virus is used to infect HighFive Tricoplusia ni cells (Invitrogen, La Jolla, CA).

Recombinantly produced proteins are isolated and purified using a variety of standard techniques. The actual techniques used vary depending upon the host organism used, whether the protein is designed for secretion, and other such factors. Such techniques are well known to the skilled artisan (see, e.g. chapter 16 of Ausubel, F. et al., "Current Protocols in Molecular Biology", pub. by John Wiley & Sons, Inc. (1994).

IV. Assays For Characterizing The Proteins

Recombinantly produced proteins are useful for a variety of purposes. For example, they can be used in in vitro assays to screen known insecticidal chemicals whose target has not been identified to determine if they inhibit protein activity. Such in vitro assays may also be used as more general screens to identify chemicals that inhibit such protein activity and that are therefore novel insecticide candidates. Recombinantly produced proteins may also be used to elucidate the complex structure of these molecules and to further characterize their association with known inhibitors in order to rationally design new inhibitory insecticides. Alternatively, the recombinant protein can be used to isolate antibodies or peptides that modulate the activity and are useful in transgenic solutions.

V. *In vivo* Inhibitor Assay: Discovery of Small Molecule Ligands That Interact with Proteins Of Unknown Function.

Having identified a protein as a potential insecticide target based on its essentiality for insect viability, a next step is to develop an assay that allows screening large numbers of chemicals to determine which ones interact with the protein. Although it is straightforward to develop assays for proteins of known function, developing assays with proteins of unknown functions can be more difficult.

To address this issue, novel technologies are used that can detect interactions between a protein and a ligand without knowing the biological function of the protein. A short description of three methods is presented, including fluorescence correlation spectroscopy, surface-enhanced laser desorption/ionization, and biacore technologies. In addition to those descibed here, there are additional methods that are currently being developed that are also amenable to automated, large-scale screening.

Fluorescence Correlation Spectroscopy (FCS) theory was developed in 1972 but it is only in recent years that the technology to perform FCS became available (Madge et al. (1972) *Phys.*

Rev. Lett., 29: 705-708; Maiti et al. (1997) *Proc. Natl. Acad. Sci. USA*, 94: 11753-11757). FCS measures the average diffusion rate of a fluorescent molecule within a small sample volume. The sample size can be as low as 10^3 fluorescent molecules and the sample volume as low as the cytoplasm of a single bacterium. The diffusion rate is a function of the mass of the molecule and decreases as the mass increases. FCS can therefore be applied to protein-ligand interaction analysis by measuring the change in mass and therefore in diffusion rate of a molecule upon binding. In a typical experiment, the target to be analyzed is expressed as a recombinant protein with a sequence tag, such as a poly-histidine sequence, inserted at the N- or C-terminus. The expression takes place in *E. coli*, yeast or insect cells. The protein is purified by chromatography. For example, the poly-histidine tag can be used to bind the expressed protein to a metal chelate column such as Ni2+ chelated on iminodiacetic acid agarose. The protein is then labeled with a fluorescent tag such as carboxytetramethylrhodamine or BODIPY® (Molecular Probes, Eugene, OR). The protein is then exposed in solution to the potential ligand, and its diffusion rate is determined by FCS using instrumentation available from Carl Zeiss, Inc. (Thornwood, NY). Ligand binding is determined by changes in the diffusion rate of the protein.

Surface-Enhanced Laser Desorption/Ionization (SELDI) was invented by Hutchens and Yip during the late 1980's (Hutchens and Yip (1993) *Rapid Commun. Mass Spectrom.* 7: 576-580). When coupled to a time-of-flight mass spectrometer (TOF), SELDI provides means to rapidly analyze molecules retained on a chip. It can be applied to ligand-protein interaction analysis by covalently binding the target protein on the chip and analyze by MS the small molecules that bind to this protein (Worrall et al. (1998) *Anal. Biochem.* 70: 750-756). In a typical experiment, the target to be analyzed is expressed as described for FCS. The purified protein is then used in the assay without further preparation. It is bound to the SELDI chip either by utilizing the poly-histidine tag or by other interaction such as ion exchange or hydrophobic interaction. The chip thus prepared is then exposed to the potential ligand via, for example, a delivery system able to pipet the ligands in a sequential manner (autosampler). The chip is then submitted to washes of increasing stringency, for example a series of washes with buffer solutions containing an increasing ionic strength. After each wash, the bound material is analyzed by submitting the chip to SELDI-TOF. Ligands that specifically bind the target will be identified by the stringency of the wash needed to elute them.

Biacore relies on changes in the refractive index at the surface layer upon binding of a ligand to a protein immobilized on the layer. In this system, a collection of small ligands is injected sequentially in a 2-5 microlitre cell with the immobilized protein. Binding is detected by surface plasmon resonance (SPR) by recording laser light refracting from the surface. In general, the refractive index change for a given change of mass concentration at the surface layer is practically the same for all proteins and peptides, allowing a single method to be applicable for any protein (Liedberg et al. (1983) *Sensors Actuators* 4: 299-304; Malmquist (1993) *Nature* 361: 186-187). In a typical experiment, the target to be analyzed is expressed as described for FCS. The purified protein is then used in the assay without further preparation. It is bound to the Biacore chip either by utilizing the poly-histidine tag or by other interaction such as ion exchange or hydrophobic interaction. The chip thus prepared is then exposed to the potential ligand via the delivery system incorporated in the instruments sold by Biacore (Uppsala, Sweden) to pipet the ligands in a sequential manner (autosampler). The SPR signal on the chip is recorded and changes in the refractive index indicate an interaction between the immobilized target and the ligand. Analysis of the signal kinetics on rate and off rate allows the discrimination between non-specific and specific interaction.

The compounds that are active in the methods disclosed herein may be used to combat agricultural pests such as aphids, locusts, spider mites, and boll weavils as well as such insect pests which attack stored grains and against immature stages of insects living on plant tissue. The compounds are also useful as a nematodicide for the control of agriculturally important soil nematodes and plant parasites.

VI. Production of peptides

Phage particles displaying diverse peptide libraries permits rapid library construction, affinity selection, amplification and selection of ligands directed against an essential protein (H.B. Lowman, *Annu. Rev. Biophys. Biomol. Struct.* 26, 401-424 (1997)). Structural analysis of these selectants can provide new information about ligand-target molecule interactions and then in the process also provide a novel molecule that can enable the development of new insecticides based upon these peptides as leads.

The invention will be further described by reference to the following detailed examples. These examples are provided for purposes of illustration only, and are not intended to be limiting

unless otherwise specified.

EXAMPLES

Standard recombinant DNA and molecular cloning techniques used here are well known in the art and are described by Sambrook, *et al.*, *Molecular Cloning*, eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989) and by T.J. Silhavy, M.L. Berman, and L.W. Enquist, *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984) and by Ausubel, F.M. *et al., Current Protocols in Molecular Biology*, pub. by Greene Publishing Assoc. and Wiley-Interscience (1987). Well known *Drosophila* molecular genetics techniques can be found, for example, in Robert, D.B., *Drosophila, A Practical Approach* (IRL Press, Washington, DC, 1986).

Example 1: Identification Of Lethal Lines

Essential nucleotide sequences are identified through the isolation of lethal mutants defective in development. The genetic scheme for mobilization of P-lacW is as performed in Deak *et. al*, *Genetics* 147: 1697-1722 (1997). Additional lethal lines are identified and disclosed in Braun, A., B. Lemaitre, *et al., Genetics* 147: 623-634 (1997); Galloni, M. and B. A. Edgar, *Development* 126: 2365-2375 (1999); Gateff, E., *Int. J. Dev. Biol.* 38(4): 565-590 (1994); Mechler, B. M. J. Biosci., *Bangalore* 19(5): 537-556 (1994); Roch, F., F. Serras, *et al., Mol. Gen. Genet.* 257: 103-112 (1998); Russell, M. A., L. Ostafichuk, *et al., Genome* 41: 7-13 (1998); and in Torok, T., G. Tick, *et al. Genetics* 135: 71-80 (1993), Schaefer et al., 1999.8.12 Personal communication to FlyBase. Furthermore, the BDGP gene disruption project of single P-element insertions reveals lethal lines mutating 25% of vital *Drosophila* genes Spradling, A. C., D. Stern, *et al., Genetics* 153: 135-177 (1999).

Males carrying the transposase source P(Δ2-3) are crossed en masse to yellow white females homozygous for a P-lacW insertion on the X chromosome. Males carrying the PlacW insertion on the X and Δ2-3 on the third chromosome are collected from this cross. The F0 "jumpstart" males are crossed in groups of 10-15 to 20-25 females of w spl; Sb/TM3, Ser genetype. Male F1 progeny with pigmented eyes indicate that the P-lacW has jumped to an autosome. An average of 10-15 males from each F0 cross lacking Δ2-3 are crossed individually to y w; DTS-4/TM3, Sb Ser females, that all third chromosomal insertions result in balanced F2 stocks. Insertions on other autosomes yield white-eyed flies in the F2 generation and are

eliminated. The balanced third chromosome insertions are tested for lethality in the next generation by placing four to six pairs of y w; P-lacW/TM3, Sb Ser flies in a vial and examining their progeny for the presence of homozygous P-lacW flies. To analyze the lethal phase, the TM3, Sb Ser balancer is replaced by the TM6C, TB Sb chromosome. In such a genetic background, homozygous mutants can be identified by their wild-type body-length. An average of 10-15 pairs of flies are placed in vials supplemented with yeast paste, and the eggs are collected from each line for 1 day. The development of 50-100 progeny is monitored, and the presence of homozygotes are recorded in all developmental stages. Lethal phase is assigned to a developmental stage in which homozygote animals last appear. Lethal lines are identified and maintained.

Table 3 P-element location

SEQ ID	Inventor's reference	p-element line	Inverse PCR	df cross
1	GIN00418, CT41283	EP(3)0831	public	Df(3L)Ar14-8
3	GIN00831, CT1581	EP(3)3137	public	Df(3R)Tpl10
5	GIN00996, CT4870	EP(2)2475	public	Df(2R)ST1
7	GIN01641, CT4036	EP(3)3522	public	Df(3R)Dr-rv1
11	GIN05114, CT35627	EP(3)1005	public	Df(3R)L127
19	GIN08522, CT13682	EP(3)0745	public	Df(3L)iro-2
21	GIN08754, CT14494	EP(3)3247	public	Df(3L)AC1
23	GIN09345, CT16487	EP(3)3343	public	Df(3L)st-f13
27	GIN09658, CT17430	EP(2)0682	public	Df(2L)prd1.7
33	GIN10694, CT20945	EP(2)2403	public	Df(2L)J39
35	GIN10918, CT21672	EP(3)3504	public	Tp(3;Y)ry506-85C
41	GIN11589, CT23419	EP(3)0572	Public	Df(3L)BK10
43	GIN11844, CT24166	EP(3)3112	Public	Df(3R)Cha7

Example 2: Sequence Determination

Inverse PCR: To determine the flanking sequence of the lethal lines, the "Inverse PCR and Cycle Sequencing Protocol for Recovery of Sequences Flanking PZ, PlacW, and PEP elements" of E. Jay Rehm, Berkeley *Drosophila* Genome Project on the world wide web at fruitfly.org/methods/ is used with slight modifications. These modifications include the following: genomic DNA is obtained from 10 flies, rather than 30 flies, with adjustments for final concentrations; all DNA precipitations are performed using glycogen; for some reactions, all of the digest volume is used in the appropriate ligations; the number of cycles in PCR

reactions was increased to 40; Pry1 and Pry2 were used to sequence the PEP line flanking sequences.

Genomic DNA isolation: Flies are collected and frozen at –20°C until ready for use. Genomic DNA is prepared by grinding flies in 200 µl Buffer A with a disposable grinder 30X (Buffer A is composed of 100 mM Tris-Cl, pH7.5, 100 mM EDTA, 100 mM NaCl, 0.5% SDS). Add 200 µl additional Buffer A; grind another 15X. Keep on ice until finished. Incubate at 65°C for 30 minutes. Vortex to mix. Add 800 µl freshly made LiCl/KAc Solution (LiCl/ Kac Solution is comprised of 1 part 5 M KAc and 2.5 parts 6 M LiCl). Vortex. Incubate –20°C for 20 minutes. Spin at maximum speed at room temperature 15+ minutes. Transfer 1 ml supernatant to a clean tube avoiding floating debris. Add 600 µl room temperature isopropanol to supernatant. Mix well by tipping. Add 0.5 µl glycogen. Vortex. Incubate at room temperature for 5 minutes. Spin 15 minutes at room temperature, maximum speed. Aspirate away the supernatant. Wash 2X with 500 µl 70% room temperature ethanol; vortex between washes. Spin for 10 minutes at room temperature, maximum speed. Aspirate away supernatant. Dry in a speed vacuum for 10 minutes. Resuspend in 50 µl TE + 0.1 mg/ml RNAse A {for 1 ml TE/RNAse A Solution, add 990 µl TE + 10 µl RNAse A (10mg/ml)). Check 5 µl on 0.8% gel.

Digest Genomic DNA (Sau3A I, HinP1 I, or Msp I--done separately): Set up digests in 96 well tray. Per reaction, add 10 µl genomic DNA, 5 µl 10X Buffer, 2 µl 0.1mg/ml RNAase A stock, 30.5 µl dH_2O, 10 units of enzyme (8 units for Sau 3A I), 0.5µl of 100X BSA (for Sau 3AI only). Incubate at 37°C for 2.5 hours. Check on 0.8% gel before heat-inactivating at 65°C for 20 minutes.

Ligate P Element and Flanking DNA: Set-up ligation tube with 400 µl of ligation mixture then add 30-50 µl of the digest: Per reaction, add 30 µl of digested genomic DNA, 43 µl of 10X ligation buffer (NEB), 375 µl of dH_2O, and 2 µl of ligase (2 Weiss units). Incubate overnight at 4°C. Total reaction volume is adjusted as appropriate.

Precipitate Ligated DNA: To ligation tube, add 40 µl 3M NaAc pH5.2 + 1ml 100% room temperature ethanol + 1 µl glycogen. Mix by tipping. Incubate –20°C for 15+ minutes. Spin 15 minutes, 4°C. Aspirate away supernatant. Wash with 500 µl room temperature 70% ethanol. Vortex. Spin room at temperature for 10 minutes. Aspirate away supernatant. Dry in speed vacuum for 10 minutes. Resuspend in 50 µl TE. Vortex to mix. Transfer to 96 well plate.

PCR: Set up PCR reactions in 96 well plates (Applied Biosystems). Set up PCR reactions with primers appropriate for the type of P element and the end of the element from which genomic sequence is to be recovered.

Primers for PCR: (type of P element 5' or 3' end forward primer reverse primer annealing temperature):

PZ P-element5' endPlac4Plac1	60°
PZ P-element3' endPry4Pry1	55°
PZ P-element3' endPry2Pry1	60°
PlacW P-element5' endPlac4Plac1	60°
PlacW P-element3' endPry4Plw3-1	55°
PlacW P-element3' endPry2Pry1	60°
PEP P-element5' endPwht1Plac1	60°
PEP P-element3' endPry4Pry1	55°
PEP P-element3' endPry2Pry1	60°

The Pry2/Pry1 combination has a higher annealing temperature than the Pry4/Pry1 and Pry4/Plw3-1 combinations, but the resulting PCR products do not allow sequencing directly off the 3' end of the P-element. The latter primer combinations are therefore used in all initial experiments; the Pry2/Pry1 combination can be used in those cases where strong and unique bands do not result.

Per reaction: 10 µl of ligated genomic DNA, 1 µl of 10mM dNTP mix, 1 µl of 10 µM forward primer stock, 1µl of 10 µM reverse primer stock, 5 µl of 10X Qiagen Taq buffer, 31.5 µl of dH_2O, 0.5 µl of Qiagen Taq.

Cycles: 1X 95°C for 5 minutes; 40X (95°C for 30 seconds; 60°C (high temp) or 55°C (low temp) for 30 seconds; 68°C for 2 minutes); 1X 72°C for 10 minutes; hold at 4°C; run 10µl on 1.5% gel to check. Rearray positive wells to 96 well plate for sequencing clean-up. The primer sets for PCR are as shown in the table below:

Table 4 PCR Primers

Digest, End, Temperature	Forward PCR Primer	Reverse PCR Primer
H5h	Plac4	Plac1
H3h	Pry2	Pry1

H3l	Pry4	Plw3-1
M5h	Plac4	Placl
M3h	Pry2	Pryl
M3l	Pry4	Plw3-1
S5h	Plac4	Placl
S3h	Pry2	Pryl
S3l	Pry4	Plw3-1

PCR Primer Sequences (5' to 3'):

Plac4 (27)	- act gtg cgt tag gtc ctg ttc att gtt	SEQ ID NO:51
Placl (24)	- cac cca agg ctc tgc tcc cac aat	SEQ ID NO:52
Pry4 (23)	- caa tca tat cgc tgt ctc act ca	SEQ ID NO:53
Pryl (26)	- cct tag cat gtc cgt ggg gtt tga at	SEQ ID NO:54
Pry2 (28)	- ctt gcc gac ggg acc acc tta tgt tat t	SEQ ID NO:55
Plw3-1 (19)	- tgt cgg cgt cat caa ctc c	SEQ ID NO:56
Pwhtl (19)	- gta acg cta atc act ccg aac agg tca ca	SEQ ID NO:57

Enzymatic Clean-Up for Sequencing: To 40 µl PCR reaction, add 4 µl of enzyme mix. Incubate at 37°C for 1 hour. Inactivate at 70°C for 10 minutes. (Enzyme Mix consists of 2.5U/µl Exonuclease I (Amersham E700732), 0.5U/µl Shrimp Alkaline Phosphatase (Amersham E70183), 1X Amplitaq PCR buffer, add dH_20 to final volume.)

Example 3: Sequence Analysis

Sequence of the flanking sequence generated by inverse PCR is performed on an ABI 3700 sequencer (Perkin Elmer) using BIG DYE sequencing reaction.

Primer sets for sequencing are as shown in the table below:

Table 5 PCR Primers for Flanking Sequences

Digest, End, Temperature	Forward Primer	Reverse Primer
H5h	Splac2	Sp1
H3h	Pry2	Sp5
H3l	Spep1	Sp5

M5h	Splac2	Sp1
M3h	Pry2	Sp5
M3l	Spep1	Sp5
S5h	Splac2	Sp1
S3h	Pry2	Sp6
S3l	Spep1	Sp6

The following primer sets are designed to sequence both ends of PCR products recovered from PlacW and PZ strains:

Splac2 and Sp1 - for use with the Plac4/Plac1 5' PCR primer combination with either PZ or PlacW P-elements; allows sequencing of both ends of the PCR fragment.

Spep1 and Sp3 - for use with the Pry4/Pry1 3' PCR primer combination with PZ P-elements; allows sequencing of both ends of the PCR fragment.

Spep1 and Sp6 - for use with the Pry4/Plw3-1 3' PCR primer combination with PlacW P-elements where Sau3a digestion is performed; allows sequencing of both ends of the PCR fragment.

Spep1 and Sp5 - for use with the Pry4/Plw3-1 3' PCR primer combination where HinP1 digestion is performed; allows sequencing of both ends of the PCR fragment.

Pry1 and Pry2 – for use with the Pry1/Pry2 3' PCR primer combination; allows sequencing of both ends of the PCR fragment.

The PCR products recovered from PEP strains are sequenced with the following primers: Sp1- for use with the Pwht1/Plac1 5' PCR primer combination with the PEP element; Spep1- for use with the Pry4/Pry1 3' PCR primer combination with the PEP element; Pry1 and Pry2 for use with the Pry1/Pry2 3' PCR primer combination with the PEP element.

Primer Sequences (5' to 3'):

Splac2 (25)	- gaa ttc act ggc cgt cgt ttt aca a	SEQ ID NO:58
Sp1 (22)	- aca caa cct ttc ctc tca aca a	SEQ ID NO:59
Sp3 (24)	- gag tac gca aag ctt taa cta tgt	SEQ ID NO:60
Sp6 (23)	- tga cca cat cca aac atc ctc tt	SEQ ID NO:61
Sp5 (25)	- gca tca caa aaa tcg acg ctc aag t	SEQ ID NO:62
Spep1 (19)	- gac act cag aat act att c	SEQ ID NO:63

Melting temperatures of sequencing primers:

Splac2- 60.1°C

Sp1- 50.6°C

Sp3- 49.3°C

Sp6- 54.9°C

Sp5 -60.3°C

Spep1- 44.8°C

Example 4: Secondary Confirmation of Lethality

The lethality of the chromosome carrying the P-element insertion is demonstrated genetically as described in Example 1. The essential *Drosophila* nucleotide sequences are identified by isolating nucleotide sequences flanking the P-element insertion and aligning those sequences with genomic *Drosophila* sequence obtained from the Celera *Drosophila* database. However, in some instances, a second site mutation exists on the chromosome that is responsible for the lethality. In other instances, the location of the flanking sequence is such that determination of which gene(s) are affected by the P-element insertion is rendered difficult or impossible. Thus, to provide secondary confirmation that the gene indicated is essential, there are many methods that one skilled in the art can use, e.g., rescue of the lethality using transformation technology, perturbation of the gene in a targeted manner, or failure to complement a deficiency.

To provide secondary confirmation, lethal lines are crossed to a line containing a deficiency. This creates a hemizygous condition in that particular region and reveals the recessive phenotype of the P-element. Complementation with deficiencies that unequivocally remove the P-element insertion site is taken as proof that the P-element does not cause the associated phenotype. Failure to complement indicates that the strain is verified. This method is as performed in Spradling, A. C., D. Stern, *et al., Genetics* 153: 135-177 (1999). If the insert is present on the X chromosome, which is present in two copies in females but only one copy in males, then the recessive phenotype of the P-element insert is revealed by this hemizygous condition in males. A rescue cross is performed to a stock containing a duplication spanning the region of the insert on the X chromosome on one of the autosomes. If the males survive then the

presence of an essential gene disrupted by the P-element but rescued by the duplication is confirmed. While lines with secondary mutations closely linked to the P insertion might be erroneously verified by these procedures, further molecular and genetic analyses suggest that the frequency of such errors is small. RNA interference, described in Fire, A., S. Xu, *et al., Nature* 391, 806-811 (1998) and Kennerdell, J.R. and Carthew, R.W., *Cell* 95, 1017-1026 (1998), is used as a method to target a gene of interest and demonstrate that the perturbation of the identified gene produces a lethal phenotype.

Example 4a: Double-Stranded RNA Interference

Preparation of dsRNA for Injection. Sequences to be expressed as dsRNA were cloned into Bluescript KS(+) (Stratagene of La Jolla, California), linearized with the appropriate restriction enzymes, and transcribed *in vitro* with the Ambion T3 and T7 Megascript kits following the manufacturer's instructions (Ambion Inc. of Austin, Texas). Transcripts were annealed in injection buffer (0.1mM $NaPO_4$ pH 7.8, 5mM KCl) after heating to 85°C and cooling to room temperature over a 1- to 24-hr period. All annealed transcripts were analyzed on agarose gels with DNA markers to confirm the size of the annealed RNA and quantitated as described previously (Fire et al. (1998) *Nature* 391(6669):806-811). Injected RNA was not gel-purified. Injection of 0.1 nl of a 0.1- to 1.0-mg/ml solution of a 1-kb dsRNA corresponds to roughly 10^7 molecules/injection.

Injection of *Drosophila melanogaster* Embryos. Fly cages were set up using 2- to 4-day flies. Agar-grape juice plates were replaced every hour to synchronize the egg collection for 1-2 days. The eggs were collected over a 30- to 60-min period for subsequent injection. The eggs were washed into a nylon mesh basket with tap water. The chorion was removed by brief soaking in a dilute bleach solution. Eggs were positioned on a glass slide such that each egg was in a same orientation. Double-stranded RNA was injected into middle of each egg using an Eppendorf transjector (Eppendorf Scientific, Inc. of Westbury, New York). Following injection, slides were stored in a moist chamber to prevent dessication of the embryos. Embryos were monitored for development and transferred as first instar larvae to vials containing *Drosophila* medium. Methods for rearing *Drosophila* staging and common genetic techniques can be found, for example, in Roberts (1986) *Drosophila melanogaster*, A Practical Approach, IRL Press, Washington, DC; Ashburner (1989a) *Drosophila*: A Laboratory Handbook, Cold Spring Harbor

Laboratory Press, New York, New York; Ashburner (1989b) *Drosophila*: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York, New York; Goldstein & Fyrberg, eds (1994) in Methods in Cell Biology, Vol. 44, Academic Press, San Diego, California.

The data in Table 6 demonstrates the lethal effect of disrupting the production of protein from the message of the specified gene through RNAi. Based on data from postitve and negative controls, a reduction in survival (%viable adults from developed eggs) below 38% represents a significant lethal effect. Many genes show a complete loss of survivability (with 0% viable). Others show a range of phenotypic penetrance, which is most likely due to the variability of the RNAi technique, but are still considered lethals because they are significantly below controls.

Table 6 Data for dsRNA Interference

SEQ ID	Inventor's reference	# eggs injected	# eggs showing morpho-logical development	# hatched larvae	# pupae	# adults	% viable adults from developed eggs
	none, buffer only	941	806	580	500	433	53.7
9	GIN02024, CT27956	87	74	60	1	1	1.4
13	GIN05842, CT4886	49	41	0	0	0	0.0
15	GIN06014, CT32725	54	48	28	1	0	0.0
17	GIN08020, CT11825	160	81	39	29	23	28.4
25	GIN09460, CT16853	77	76	63	6	1	1.3
29	GIN10467, CT20131	163	143	106	37	21	14.7
31	GIN10517, CT20377	85	82	41	33	31	37.8
37	GIN11550, CT20832	72	64	47	4	0	0.0
39	GIN11578, CT23580	85	80	65	30	27	33.8
45	GIN11932, CT20784	104	92	67	45	29	31.5
47	GIN12213, CT24821	75	68	43	0	0	0.0
49	GIN12858, CT26398	72	52	0	0	0	0.0

Example 5: Isolation Of Full Length cDNA

A cDNA screen is performed using a *Drosophila melanogaster* cDNA library probed with a portion of each nucleotide sequence disclosed in the Sequence Listing. Positive colonies are selected, a subset sequenced, and a clone corresponding to the full-length cDNA is recovered. Alternatively, primers from the predicted 5' and 3' end are used in polymerase chain reaction with either a *Drosophila* cDNA library or first strand cDNAs obtained by reverse transcription of *Drosophila* mRNAs as template to amplify a fragment representing the full-length clone.

Example 6: Expression Of Recombinant Protein In Insect Cells

Baculovirus vectors, which are derived from the genome of AcNPV virus, are designed to provide high levels of expression of cDNA in the SF9 line of insect cells (ATCC CRL# 1711). Recombinant baculovirus expressing the cDNA of the present invention is produced by the following standard methods (InVitrogen MaxBac Manual): cDNA constructs are ligated into the polyhedrin gene in a variety of baclovirus transfer vectors, including the pAC360 and the BleBAc vector (InVitrogen). Recombinant baculoviruses are generated by homologous recombination following co-transfection of the baculovirus transfer vector and linearized AcNPV genomic DNA (Kitts, P.A., *Nucleic Acid. Res.* 18: 5667 (1990)) into SF9 cells. Recombinant pAC360 viruses are identified by the absence of inclusion bodies in infected cells and recombinant pBlueBac viruses are identified on the basis of B-galactosidase expression (Summers, M.D. and Smith, G.E., Texas Agriculture Exp. Station Bulletin No. 1555). Following plaque purification, the *Drosophila* cDNA expression is measured.

The cDNA encoding the entire open reading frame for the *Drosophila* cDNA is inserted into the BamHI site of pBlueBacII. Constucts in the positive orientation, which are identified by sequence analysis, are used to transfect SF9 cells in the presence of linear AcNPV wild type DNA. Authentic, active *Drosophila* cDNA is found in the cytoplasm of infected cells. Active *Drosophila* cDNA is extracted from infected cells by hypotonic or detergent lysis.

Example 7: Expression Of Recombinant Protein In *E. coli*

A cDNA clone of the present invention is subcloned into an appropriate expression vector and transformed into *E. coli* using the manufacturer's conditions. Specific examples include plasmids such as pBluescript (Stratagene, La Jolla, CA), pFLAG (International Biotechnologies,

Inc., New Haven, CT), and pTrcHis (Invitrogen, La Jolla, CA). *E. coli* is cultured, and expression of the recombinant protein is confirmed. Recombinant protein is then isolated using standard techniques.

Example 8: *In vitro* Binding Assays

Recombinant protein is obtained, for example according to Example 6 or Example 7. The protein is immobilized on chips appropriate for ligand binding assays. The protein immobilized on the chip is exposed to sample compound in solution according to methods well know in the art. While the sample compound is in contact with the immobilized protein measurements capable of detecting protein-ligand interactions are conducted. Examples of such measurements are SEDLI, biacore and FCS, described above. Compounds found to bind the protein are readily discovered in this fashion and are subjected to further characterization.

The above disclosed embodiments are illustrative. This disclosure of the invention will place one skilled in the art in possession of many variations of the invention. All such obvious and foreseeable variations are intended to be encompassed by the appended claims.

The numerous publications and patents referred to in this document are hereby incorporated by reference, in their entirety.